

Kathrin Steyer (Mannheim)

Idiomatik hypermedial. Zur Repräsentation von Wortverbindungen im Informationssystem *Wissen über Wörter*

1 Einleitung

Regularitäten von Wortverbindungen bzw. Kombinationen, die Wörter in einer bestimmten sprachlichen Umgebung miteinander eingehen, werden derzeit nicht mehr nur von Idiomatikforschern analysiert und beschrieben, sondern rücken generell in den Fokus der Linguistik, und zwar in zweifacher Hinsicht: Es wächst einerseits das Interesse an Mehrworteinheiten als eigenständigem Forschungsgegenstand. Andererseits dienen syntagmatische Korpusanalysen als Methode zur Verifizierung allgemeiner sprachlicher Phänomene. Eine Vorreiterrolle nehmen Lexikografie und Fremdsprachendidaktik ein (vgl. Hausmann 1984 usw.), die die Relevanz von sprachlichen Kombinationsmodi für ein kulturell angemessenes Kommunizieren schon seit längerem erkannt haben (vgl. Stubbs 1997). Verbindungskonventionen werden folgerichtig als lexikografische Informationen bzw. als Lerninhalte aufbereitet und zur Verfügung gestellt. Auch die Pragmatik interessiert sich verstärkt für 'formelhafte Ausdrücke' als elementare Kommunikationskonstituenten und zwar unter der Prämisse, dass Sprachgebrauch in einem hohen Maße von Vorgefertigtem bzw. Vorformuliertem bestimmt ist (vgl. Gülich/Krafft 1998). Wortkombinationen indizieren demzufolge in besonderer Weise Thema-, Sorten- und Kontextschemata (Feilke 2000: 5). Die syntagmatische Analyse erfährt schließlich deshalb eine Aufwertung, weil sie die Möglichkeit eröffnet, Strukturen des Lexikons empirisch zu verifizieren und damit die Methode der reinen Introspektion zu Gunsten einer am Sprachgebrauch orientierten linguistischen Beschreibung sprachlichen Wissens aufgeben zu können (vgl. u.a. Feilke 1997, Heringer 1999). Zunehmend setzt sich dabei die Erkenntnis durch, dass man diese kookkurrenzellen Regularitäten nur bis zu einem gewissen Grade mittels der eigenen Sprachkompetenz bzw. auf der Basis tradierten Expertenwissens erklären kann (vgl. u.a. Boguraev/Pustejovsky 1996; Haß-Zumkehr 2002). Trotz dieser Einsicht befinden sich die Versuche, Mehrwortregularitäten auf einer empirisch ernstzunehmenden und damit systematischen Basis zu analysieren, noch in den Anfängen. Eine Ursache liegt ohne Zweifel in der großen Kluft zwischen der fortschreitenden Entwicklung der Korpus-technologie (verbunden mit der Entwicklung automatischer Recherchewerkzeuge, wie z.B. Kollokationstools) auf der einen Seite und der linguistischen Adaption und 'Ausfüllung' auf der anderen.

Das IDS-Projekt *Wissen über Wörter* (WiW) beschreitet sowohl hinsichtlich der Aufbereitung und Beschreibung von üblichen Wortverbindungen einer Sprache als

auch bei der Entwicklung und Anwendung der empirischen Methode der korpusbasierten Kookkurrenzanalyse als linguistisches 'Arbeits- und Denkinstrument' neue Wege. *Wissen über Wörter* ist das lexikologisch-lexikalische Informationssystem des Instituts für Deutsche Sprache Mannheim (IDS), das auf einer speziell für digitale Medien entwickelten Wörterbuchstruktur basiert und künftig als Schnittstelle zwischen lexikografischem Arbeitsplatz, Sprachdatenbank und einer im Internet abrufbaren Benutzeroberfläche fungieren soll. Hier geht es also nicht um die elektronische Aufbereitung herkömmlicher Printwörterbücher, sondern um ein neuartiges hypermediales System, das umfassende Informationen zum deutschen Wortschatz sowohl für Lexikografen als auch für Nutzer verfügbar machen wird; u.a. durch dynamische Korpusrecherchen, eine abgestufte Navigation in einem mehrdimensionalen Suchraum und ein komplexes Verlinkungssystem.¹ Die vielfach beklagten Begrenzungen der papiergebundenen Lexikografie können überwunden und linguistische Beschreibungen auf der Basis einer quantitativ aussagekräftigen Menge 'realer Sprachdaten' ermöglicht werden. Unabdingbar für ein solches Vorhaben ist eine integrative Herangehensweise, die versucht, korpusstatistische und philologische Denkweisen miteinander in Einklang zu bringen.

Ein Kernkonzept bilden in diesem Kontext usuelle Wortverbindungen. Darunter verstehen wir alle jene Wortverbindungen, die eine historisch gewachsene Gebrauchsnorm repräsentieren, wobei sich der usuelle Status aus der statistischen Signifikanz der Wortverbindung – dem Grad der lexikalischen Kohäsion zwischen zwei Wortformen – im Korpus ergibt. Wir sprechen also von einer korpusbezogenen Usualität.² Durch die Korpusanalyse können auch jene Mehrworteinheiten erfasst werden, die nicht zum engen Bereich idiomatisch-phraseologischer Ausdrücke gehören und die durch ihre strukturelle Unauffälligkeit oft nicht als solche erkannt werden. Aber gerade diese Ausdrücke weisen einen hohen Grad an Formelhaftigkeit auf; gerade über sie konstituiert sich Sprachgebrauch in nicht unerheblichem Maße. Bei den usuellen Wortverbindungen unterscheiden wir zwischen einzelwortbezogenen Kookkurrenzen (im Kern binäre Relationen = Kollokationen) und 'echten' Mehrwortverbindungen (u.a. Phraseologismen, Routineformeln, Funktionsverbgefüge usw.).³ Im Folgenden soll die Repräsentation von Mehrwortverbindungen in *Wissen über Wörter* vor allem in Bezug auf ihre Nutzbarkeit für die Linguistik im Allgemeinen und die Idiomatikforschung im Besonderen dargestellt werden.⁴

¹ Ausführliche Informationen auf der WiW-Homepage <http://www.ids-mannheim.de/wiw> und bei Fraas/Haß-Zumkehr 1998, Haß-Zumkehr 2001)

² WiW basiert auf den IDS-Korpora, die derzeit ca. 1846 Millionen Textwörter umfassen und die größte Korpusammlung deutschsprachiger Texte weltweit darstellen. Informationen dazu unter <http://www.ids-mannheim.de/kt/corpora.shtml>. Zur Problematik der Repräsentativität und der Aussagekraft der IDS-Korpora vgl. Steyer 2000: 112.

³ Zur linguistischen Begründung und klassifikatorischen Differenzierung dieses Konzepts vgl. Steyer 2000.

⁴ Die einzelwortbezogenen Kollokationsangaben können in WiW innerhalb der Angabe 'Kookkurrenz' abgerufen werden.

2 Das WiW-Mehrwortlemma – lexikografische Parameter

In WiW sind drei Lemmatypen vorgesehen, deren Mikrostruktur in allen wesentlichen Angaben übereinstimmt: Einwortlemmata, Mehrwortlemmata und Wortelementlemmata (u.a. Angaben zu Affixen wie ‘ver-’ und ‘-lich’). Vier strukturierte, lesartengebundene Informationsdimensionen sind für alle Lemmata verbindlich: Bedeutung&Verwendung; Grammatik; Geschichtliches&Sachliches; Kritisches&Normatives. Jedem Lemma werden lesartenübergreifend Angaben zur Schreibung&Aussprache, zu konzeptuellen Relationen zwischen den Lesarten (z.B. Polysemie, Metaphorisierung etc.) und zum Kookkurrenzverhalten des Lemmas zugeordnet. Wir werden auf folgende drei Aspekte des Artikeltyps ‘Mehrwortlemma’ näher eingehen: Lemmaauswahl, binnenstrukturelle Validierung und pragmatische Angaben.

2.1 Zur Auswahl der Kandidaten für einen Mehrwortartikel

Die Kandidaten für die Mehrwortartikel werden mit Hilfe der am IDS von Cyril Belica entwickelten statistischen COSMAS-Kollokationsanalyse ermittelt.⁵ Dieses Werkzeug ermöglicht nicht nur die Erfassung signifikanter einzelwortbezogener binärer kookkurrenzieller Relationen (Kollokationen), sondern auch eine differenzierte kontextsensitive Recherche. U.a. durch das statistische Verfahren ‘Autofokus’ können Muster und Ordnungsprinzipien, phrasale Auffälligkeiten bzw. Festigkeiten in der Umgebung dieser Zweiwortverbindungen analysiert werden.⁶ Diese auffälligen phrasalen Muster stellen Indikatoren für signifikante Mehrwortverbindungen dar. Mit einer besonderen Parametereinstellung lassen sich diese signifikanten Fügungen automatisch vorsortieren, was wir am Beispiel des Bezugslemmas ‘frei’ illustrieren wollen.⁷ Der erste fettgedruckte Kookkurrenzpartner ist dabei der primäre Kollokator; die folgenden Kookkurrenzpartner spezifizieren die Umgebung der jeweiligen Kollokationsrelation:

⁵ Zur Funktionsweise vgl. Belica 2001; Steyer 2003.

⁶ Das COSMAS-Kollokationstool geht damit über die meisten der bisher entwickelten Werkzeuge dieser Art hinaus. Es handelt sich im Gegensatz zu vielen anderen Tools um ein interaktives Modul, das auf beliebige Korpora aufsetzen und sich so flexibel an die jeweiligen Korpusgegebenheiten anpassen kann.

⁷ Natürlich kann auch hier in letzter Instanz nicht auf die linguistische Kompetenz verzichtet werden, da es sich in vielen Fällen um eine Art ‘verdeckter’ Usualität handelt, die erst bei Einbeziehung der Kontexte erkannt wird.

Suchwort: &frei

BelegNr	Gamma	Fokus ⁸	Kollokatoren	Häufigkeit ⁹
924+2:	2118	1;1	Fuß gesetzt Kauti on	2
926+83:			Fuß gesetzt	83
1009+9:			Fuß Kauti on	9
1018+111:			Fuß	111

Es liegt also nicht nur ein signifikantes Vorkommen der Relation frei-Fuß vor, sondern in ihrer Umgebung weist das flektierte Verb 'gesetzt' und das Nomen 'Kauti-on' eine auffällige Signifikanz auf.

= Mehrwortlemma: [(gegen Kauti on) auf freien Fuß setzen]

Weitere Beispiele:

BelegNr	Gamma	Fokus	Kollokatoren	Häufigkeit
1909+10:	797	1;1	Fahrt Bürger	10
1919+91:			Fahrt	91

= Mehrwortlemma: [Freie Fahrt (für freie Bürger)]

BelegNr	Gamma	Fokus	Kollokatoren	Häufigkeit
2225+2:	678	1;1	Lauf lassen können Phanta- sie	2
2227+3:			Lauf lassen können	3
2230+4:			Lauf lassen Phantasie	4
2234+19:			Lauf lassen	19
2253+1:			Lauf gelassen Phantasie	8
2254+8:			Lauf gelassen	1
2262+1:			Lauf Phantasie	8

= Mehrwortlemmata: [seiner Phantasie freien Lauf lassen]
[etw. seinen freien Lauf lassen]

⁸ Mit 'Fokus' wird die typische Kollokationsspanne angegeben; 1;1 bedeutet also, dass der Kollokationspartner typischerweise direkt vor bzw. nach dem Bezugswort auftritt. Kollokationen mit einem solch engen Grad syntaktischer Gebundenheit weisen eine ausgeprägte Modifikationsresistenz auf.

⁹ Diese Häufigkeitsangabe bezieht sich auf die Treffermenge der statistischen Zufallsauswahl; auf die gesamte Trefferzahl bezogen potenziert sie sich entsprechend.

BelegNr	Gamma	Fokus	Kollokatoren	Häufigkeit
5508+7:	66	1;1	Spiel Kräfte	7
5515+26:			Spiel	26

= Mehrwortlemma: [freies Spiel (der Kräfte)]

Weitere Kandidaten für einen Mehrwortartikel des Bezugslemmas 'frei' wären z.B.:

Himmel	[unter freiem Himmel]
Markt	[freier Markt]
Meinungsäußerung	[Recht auf freie Meinungsäußerung]
Wildbahn	[freie Wildbahn]
erfunden	[frei erfunden]
völlig	[völlig frei] ('Der Fußballer stand völlig frei vor dem Tor und schoss daneben')]
Hand	[jmdm. freie Hand lassen]
Fall	[freier Fall]
Stücken	[aus freien Stücken]
Geleit	[freies Geleit]
Rede	[freie Rede]
frank	[frank und frei]
empfangbar	[frei empfangbar] ('frei empfangbares Fernsehen')
verfügen	[frei verfügen]
Radikale	[freie Radikale]

Auch für den klassischen phraseologischen Bereich erbringt unsere automatische Kollokationsanalyse einen quantitativ großen Ertrag. So sind in den Listen für das Bezugslemma 'Kopf' über 50 idiomatische Wendungen zu finden, die zum großen Teil bereits lexikografisch kodifiziert sind, aber auch solche 'aktuelleren' wie: 'Mauer' [Mauer in den Köpfen], 'Fisch' [Der Fisch stinkt vom Kopf (her)] oder 'Herz' [mit Kopf und Herz].

2.2 Binnenstrukturelle Validierung – Normalform, typische Verwendungsmuster und korpusbasierte Modifikationen

Die idealisierten Normalformen (Nennformen) lexikalischer Einheiten – z.B. als Basis für einen Wörterbucheintrag – unterscheiden sich oft beträchtlich. So finden sich Formen, bei denen die interne bzw. externe Valenz eingeflossen ist wie 'jm. Steine in den

Weg legen'¹⁰ (*Schemann*), Formen, die Phrasen ohne Valenz angeben wie 'schwer von Kapee sein' (*Duden11*) oder auch Formen, die einen ganzen Satz darstellen wie 'da liegt der Hund begraben' (*DuW*) usw. Diese Unterschiedlichkeit hat u.E. auch weitreichende Konsequenzen für die Analyse und Beschreibung von okkasionellen vs. usuellen Variationen und Modifikationen sowie von morphosyntaktischen Restriktionen und Defekten (vgl. dazu Burger 1998: 176f.). Mit einem so grundlegenden Arbeitsprinzip wie dem der strengen Korpusbasiertheit ist man nun nicht mehr auf das mitunter täuschende Kriterium der eigenen Sprachintuition angewiesen. Die statistische Kookkurrenzanalyse hilft, der invarianten semantischen bzw. semantisch-syntaktischen Struktur einer Mehrworteinheit näher zu kommen und somit den obligatorischen Teil zu bestimmen. Es geht also immer um eine Invarianz in Relation zum Korpusvorkommen der Mehrworteinheit. Die entsprechenden Muster lassen sich aus den KWIC-Zeilen der Suchwörter bzw. der Kookkurrenzlisten ableiten:

Ausschnitt aus der Kwic-Übersicht 'Schrot und Korn'

B97 wohl, dass Manni Bender, ein Mann von **Schrot und Korn**, nach der Niederlage

B98 rd Girndt. Ein Handwerker von altem **Schrot und Korn**, der seine [...]

M89 aus Rosenheim sind Bayern von echtem **Schrot und Korn**. Vor allem aber sind sie

Einen Kapellmeister von echtem **Schrot und Korn** hatte die Musikalische A
, Strube ist sicherlich von dem rechten **Schrot und Korn**, das die BASF in den kom

T88 und Eckel noch von richtigem **Schrot und Korn** waren, ihre Fußball-Enke

T89 Ausbilder, ein Herr K. von älterem **Schrot und Korn**. Verrät er mir. Gabriele

T90 Von ganz anderem **Schrot und Korn** ist da Dr. Robert.

T91 konntest, ein echter Deutscher aus **Schrot und Korn**, bei dem ein Handschlag

Hier liegt eine syntaktische Invarianz (Nebenordnung) vor. Auf lexikalischer Ebene gibt es invariante Elemente ('Schrot', 'Korn', 'und') und variable. Präpositionen ('von', 'aus') und Adjektive ('echtem', 'altem' usw.), die im Regelfall als zur Normalform zugehörig interpretiert werden, sind austauschbar. Die folgenden Beispiele sollen die Vielfalt von aus den Korpusbelegen extrahierten invarianten Strukturen verdeutlichen:

[Friede, Freude, NOMEN]
[wilde Ehe]
[wohin die Reise geht]
[an allen Ecken und Enden]
[Eigenlob stinkt]
[Stuhl vor die Tür]
[Schrot und Korn]

¹⁰ Solche Formen finden in WiW bei den grammatischen Phrasenmusterangaben, den Angaben zur syntaktischen Binnenstruktur bzw. in den Angaben zu semantischen Rollen Berücksichtigung.

[Stein des Anstoßes]
 [NEG+hinterm Ofen hervorlocken]
 [Steter Tropfen höhlt den Stein]
 [dünn besiedelt]
 [wo Barthel den Most holt]
 [in der Schwebe]
 [Fell des Bären verteilen]
 [wie eine Bombe einschlagen]
 [Korb+VERB]

Diese korpusbezogenen Invariantenangaben dienen in WiW als formale Basis für das Artikelstichwort. Invariantenkomponenten sind die jeweils fixen lexikalischen Einheiten und – fakultativ – metasprachliche Markierungen. Solche metasprachlichen Markierungen können u.a. Wortartenangaben (z.B. VERB) oder auch obligatorische Negationskomponenten (NEG) sein. Aus der Invariantenangabe lassen sich darüber hinaus die (autosemantischen) Basiselemente im Sinne von Fleischer (1997) ableiten, z.B. ‘Schrot’ bzw. ‘Korn’ bei ‘Schrot und Korn’. Die flektierten Formen eines Basiselements werden jeweils auf eine invariante Grundform reduziert (z.B. Nominativ), wobei es Ausnahmen gibt: Kommt ein Basiselement beispielsweise nur in ein und derselben flektierten Form im Korpus vor wie bei ‘Eigenlob *stinkt*’, dann ist diese invariant. In vielen Fällen – das zeigen schon die wenigen hier angeführten Beispiele – handelt es sich um Formen, die in ihrer teils kryptischen Ausprägung schwer rezipierbar sind. Dobrovol’skij spricht in diesem Zusammenhang vom “Widerspruch zwischen der lexikographischen Universalität der Nennform [...], die aber oft befremdlich wirke, und der psychologisch realen und damit benutzerfreundlicheren Aussageform, die aber nie alle Realisationsformen abdecken könne (1993: 59). Um diesem “Hauptdilemma der Phraseographie” (ebd.) zu entgehen, haben wir mehrere Abstraktionsstufen angesetzt und neben der Invariantenangabe noch zwei weitere Angaben eingeführt: die Angabe zu typischen Verwendungsmustern und die Angabe zu korpusbasierten Modifikationen bzw. Variationen. Die Angaben zu typischen Verwendungsmustern stellen eine mittlere Abstraktionsebene dar. Sie lassen sich oft auch als Syntagmenangaben auffassen und kommen der natürlichen Aussageform am nächsten. Sie geben – ebenfalls korpusbasiert – typische, gebräuchliche Muster bzw. Einbettungen der jeweiligen Mehrwortverbindung an, die aber noch nicht die reale Verarbeitung in Texten darstellen, sondern eine leicht idealisierte Form. Anhaltspunkte bietet auch hier wieder die Kookkurrenzanalyse, die durch die feine Analyse des Kontextes auch Aufschluss über typische Kontextpartner einer Wendung gibt:

aus echtem Schrot und Korn/von echtem Schrot und Korn/aus altem Schrot und Korn/von altem Schrot und Korn
 an allen Ecken und Enden sparen/es fehlt an allen Ecken und Enden
 den Stein der Weisen finden
 Korb geben/Korb holen/Korb bekommen

ins Reine bringen/im Reinen sein
sein Metier/Handwerk aus dem Effeff kennen/beherrschen

Ergänzt werden diese Angaben dann durch beliebig viele Korpusbelege, in denen die Mehrworteinheiten in ihrer realen Sprachumgebung aufgeführt werden, z.B.

B98/805.28473 Berliner Zeitung, 09.05.1998

Problematisch seien vor allem Psychopathen. "Ein echter Gangster von altem **Schrot und Korn** macht keine Probleme. Mit dem geht man vertrauensvoll Kaffee trinken, wenn er seine Strafe abgesessen hat."

T95/JUL.31667 die tageszeitung, 24.07.1995

Dort herrscht nach wie vor ein Riese, den viele schon längst totgesagt hatten, "Big Blue" IBM. Der Konzern mit Sitz in Armonk bei New York zeigte dem Emporkömmling von der Westküste erst kürzlich wieder, wo der Barthel den Most holt: In der IBM-Bilanz stand ein Gewinn von 4 Milliarden Mark - nur für die letzten sechs Monate.

3 Zur Art der Angaben im WiW-Mehrwortlemma – Beispiel: pragmatische Informationen

Neben den für alle Lemmatypen geltenden Angaben sind einige in besonderem Maße für Mehrwertverbindungen relevant. Im Bereich der semantischen Beschreibung betrifft dies vor allem Bedeutungsparaphrasen, die Angaben zur Motiviertheit und Idiomatizität, auf der grammatischen Ebene die Informationen zur syntaktischen Binnenstruktur, die Phrasenmusterangaben bzw. Angaben zur internen und externen Valenz und bei den diachron-enzyklopädischen Herleitungen Informationen zu Herkunftsbe- reich und zur kulturellen Symbolik. Dabei werden die Vorteile eines korpusbasierten Hypertextsystems, das keine Platzprobleme kennt und neben objektsprachlichen und listenbezogenen Angaben auch unbegrenzte textuelle Erläuterungen zulässt, mehr als deutlich. Besonders im Bereich der pragmatischen Interpretation bietet unser System Beschreibungsmöglichkeiten, die nicht mehr vergleichbar sind mit bisher praktizierten pragma-stilistischen Etikettierungen in Wörterbüchern. Wir wollen dies anhand von ausformulierten Angaben zur Diskursspezifik¹¹ bei der Verwendung einer Mehrwert- einheit (DiskursA); zu Sprechereinstellungen und –bewertungen (EinstA) und zum illokutiven Potenzial (IllokA) verdeutlichen.

¹¹ Diskurs verstehen wir in einem relationalen Sinne als Vernetzung von Texten und Äußerungen, zu- meist im Kontext des öffentlichen Diskurses (vgl. Steyer: 1997). Auch hier bietet uns unsere Ko- okkurrenzanalyse erste Indikatoren für über den Einzeltext hinausgehende thematisch-argumentative Stereotypisierungen und Bewertungen, die mit einem Wort bzw. einer Wortverbindung im Zusammenhang stehen.

Mehrwortlemma: 'Standort Deutschland'

<DiskursA> Die signifikanten verbalen Kookkurrenzpartner dieser öffentlichen Formel, z.B. 'kaputtreden', 'sichern', 'meiden', 'schaden', 'retten', 'lamentieren' usw. weisen auf die dominante Thematisierung im öffentlichen Diskurs hin: Der Standort Deutschland wird zumeist aus einer Negativperspektive betrachtet. Er wird entweder als allgemeines Desiderat thematisiert oder als gefährdeter Wirtschaftsfaktor, gleichsam als Warnung vor Verlust, kontextualisiert. **<DiskursA>**

Mehrwortlemma: 'harter Hund'

<EinstA> Die Korpusbelege indizieren zunächst in vielen Fällen eine negative Konnotation:

<Beleg> T91/MAI.21350 *die tageszeitung*, 29.05.1991

Was wird sich in der SPD-Zentrale in Bonn ändern? Sie haben ja schon angekündigt, dass Sie "ein harter Hund" sind und sich die Leute auf einiges gefasst machen müssen. **<Beleg>**

Bei der Analyse der Kontexte ist jedoch festzustellen, dass diese Bewertung argumentativ oft in eine positive eingebettet ist: [Jemand hat zwar diese Eigenschaft, die im allgemeinen Verständnis negativ bewertet wird, aber unter bestimmten Umständen sind solche Verhaltensweisen auch positiv zu bewerten.]

<Beleg> Z98/811.06782 *Die Zeit*, 19.11.1998

Ein ehemaliger Mitarbeiter der Vereinsbankgruppe urteilt gar: "Wenn es sein muss, geht der über Leichen." Unfair, gar brutal, nein diesen Vorwurf will ein Mitglied des Vorstandes nicht gelten lassen. "Harter, sehr harter Hund, einverstanden. Aber immer kollegial."

<Beleg> T97/JUN.26087 *die tageszeitung*, 17.06.1997

Corts habe den Beweis dafür antreten müssen, dass auch er ein harter Hund sein könne. So erreichte ihn denn auch ein dickes Lob von seiner Konkurrentin für seine "klare Haltung". Erika Steinbach: "Ein fähiger Mann." Ordnung muss schließlich das Credo eines Ordnungsdezernenten sein **<Beleg>**

Da diese adversative argumentative Einbettung bei fast allen Belegen festzustellen ist, kann man hier durchaus von einer diskurstypischen Zuschreibung sprechen. **<EinstA>**

Mehrwortlemma: 'Bären aufbinden'

Die Bedeutungsparaphrase lautet:

<SemParA> Jemandem einen Sachverhalt als wahr mitteilen, obwohl man sich über seine Unwahrheit im Klaren ist. **<SemParA>**

Der eigentliche kommunikative Effekt dieser Wendung wird aber erst bei der illokutiven Analyse deutlich:

<IllokA> 1. Wenn ein Sprecher diesen Phraseologismus verwendet, ist damit zumeist ein exklamativer Sprechakt (Ausruf) folgender Ausprägung:

gung verbunden: 'Der hat dir/jmd. aber einen Bären aufgebunden!'. Der Sprecher drückt damit nicht nur aus, dass er weiß, dass diese Äußerung unwahr ist, sondern er signalisiert gleichzeitig eine gewisse Freude oder zumindest Verwunderung darüber, dass der Adressat der Äußerung diese Unwahrheit nicht erkannt hat. 2. Bei einem weiteren -im Korpus sehr häufig vorkommenden - Verwendungstyp kommt eine andere Perspektive ins Spiel: Der Adressat einer unwahren Äußerung verwendet diesen Phraseologismus selbst, um zu signalisieren, dass er den Aussagegehalt der Äußerung als unwahr und auch die Intention des Äußerungsproduzenten erkannt hat und weist diese Äußerung zurück. <illo-ka>

Für Nutzer mit nicht rein fachbezogenem Nachschlagebedürfnis kann man in der dafür vorgesehenen Zusammenfassungsangabe die Bedeutung und den kommunikativen Effekt dieser Wendung noch einmal allgemeinverständlich paraphrasieren, etwa derart:

<ZusfassA> = Jemandem überzeugend etwas Unwahres erzählen, oft mit leicht heiterem Unterton in Form von Berichten oder Geschichten. Sprecher drücken mit diesem Phraseologismus auch eine gewisse Schadenfreude darüber aus, dass ein Adressat eine unwahre Äußerung geglaubt hat, dass dieser quasi darauf hereingefallen ist ('Der hat dir aber einen Bären aufgebunden!'). Sprecher können mit diesem Phraseologismus aber auch dem Urheber dieser Äußerung deutlich machen, dass sie sie nicht glauben. ('Ich lass mir doch keinen Bären aufbinden.').
<ZusfassA>

4 Nutzerperspektive

4.1 Abrufbarkeit der Mehrwortlemmata

In WiW wird es verschiedene Zugangsmöglichkeiten zu den Mehrwortartikeln bzw. differenzierte Abrufmöglichkeiten von Wortverbindungen geben, u.a. über die alphabetische WiW-Lemmaliste, die man sich als Ganzes aufrufen und durchsehen kann. Die (invarianten) Basiselemente der jeweiligen Mehrworteinheit werden dort als anklickbare Subeinträge zu den Einwortlemmata aufgeführt, z.B.

Einwortlemma:	'Kopf'
Subeintrag:	'Dach'
	'Nägel'
	'Sand'
	'Füße'
	'stecken' usw.

Parallel dazu bieten die Kookkurrenzangaben, die zu jedem Lemma vorgesehen sind, neben der Möglichkeit, nach Kookkurrenzen je aktuell online zu recherchieren, auch eine gesonderte Liste aller Kandidaten für die Mehrwortlemmata. Wenn man sie anklickt, kommt man durch einen Link direkt zum entsprechenden Mehrwortartikel. Eine weitere Möglichkeit, nach Mehrworteinheiten zu recherchieren, sind Suchanfragen – ähnlich wie in Suchmaschinen mit Booleschen Operatoren – z.B. 'Hund AND begraben'. Durch unsere differenzierten Präsentationsformen ist der Nutzer nicht auf eine bestimmte Formulierung bei der Suche nach Mehrwortverbindungen festgelegt; sie

lassen sich sowohl in ihrer Aussageform suchen (z.B. Suchanfrage: 'Da liegt der Hund begraben') als auch in Formen wie 'jmd. legt jmdm. Steine in den Weg'. Als Recherchegrundlage dienen die Normalformen des Artikelstichwortes, die typischen Verwendungsmuster und/oder auch die grammatischen Phrasenmusterangaben (u.a. Angaben zur Valenz). Der Nutzer kommt also mit sehr verschiedenen Eingaben zu einem Rechercheergebnis.

4.2 WiW-Mehrwortlemmata als Analysebasis und Informationspool für die linguistische Forschung

WiW soll nicht nur als elektronisches Nachschlagewerk für den Deutschlerner und den sprachinteressierten Laien dienen, sondern auch Informationen und empirische Daten für linguistische Forschungen liefern. In Bezug auf die Idiomatik im weitesten Sinne ist diese Forschungsbasis sehr facettenreich:

Klassifikationen. Es sind Angaben zur Klassifikation von Mehrworteinheiten vorgesehen (sogenannte Expertenkategorien), die in Form von Listen ohne Hierarchisierung angeboten werden. Durch die Möglichkeit, beliebig viele Kreuzklassifikationen und zusätzliche Spezifikationen vorzunehmen, entfällt die Entscheidung für ein bestimmtes Klassifikationsmodell, eine ja oft schwierige Entscheidung, weil Mehrworteinheiten sehr verschieden motiviert und an ganz unterschiedliche sprachliche Phänomene gebunden sind. Der Nutzer erhält also ein sehr viel differenzierteres Bild von Mehrwortklassen und kann sich seine Gruppen variabel zusammenstellen. Er kann nach ganz speziellen Klassen suchen, die in WiW aus den Korpora aufbereitet wurden, z.B. nach allen kommunikativen Formeln, Zitaten, Modellbildungen oder Funktionsverbgefügen.

Neue Mehrwortverbindungen, neue Lesarten. Ein spezielles Teilprojekt in WiW erfasst Neologismen, so dass neologische Wendungen und Phraseologismen perspektivisch fortlaufend und auf dem neuesten Stand zur Verfügung gestellt werden.¹² Durch die Korpusanalyse können auch neue Lesarten einer Mehrwortverbindung identifiziert werden, die dann in den Mehrwortartikeln zu beschreiben sind.

Binnenstrukturelle Informationen. Die Möglichkeiten der Verifizierung (binnen)struktureller Merkmale von Mehrworteinheiten durch die Korpusanalyse haben wir in 2 bereits beschrieben. Neben der Invarianten- und Mustererkennung lassen sich auch unikale Komponenten, skalare Modifikationsanfälligkeiten bzw. -resistenzen neu bestimmen und für Forschungszwecke nachvollziehen.

Thesauri und Felder. In der Idiomatikforschung gibt es eine Vielzahl von Versuchen, Mehrwortverbindungen unter verschiedensten Gesichtspunkten zu ordnen und zu gruppieren, u.a. nach formal-strukturellen, kognitiv-semantischen, ontologisch-referen-

¹² Z.B. solche Fügungen wie 'fit wie ein Turnschuh', 'den Ball flachhalten', 'in trockenen Tüchern sein'.

ziellen oder auch kommunikativ-funktionalen Kriterien.¹³ In WiW wird diesem Ordnungsaspekt auf zweierlei Weise Rechnung getragen: Auf der einen Seite sind in bestimmten Teilen der Mikrostruktur Angaben zu Gruppenzugehörigkeiten obligatorisch, z.B. bei den Listen der Mehrwortkandidaten, die zu einem Lemma gehören (vergleichbar mit phraseologischen Reihen, z.B. alle Mehrwortverbindungen mit 'Kopf') oder bei den paradigmatischen Angaben (z.B. Aufzählung von phraseologischen Synonymen). Auf der anderen Seite wird es durch unser komplexes und stark ausdifferenziertes Annotationssystem möglich sein, dass der Nutzer selbst Feldzugehörigkeiten (in allen Informationsdimensionen) recherchieren und flexibel zusammenstellen kann. Dies sei nur kurz für einige Aspekte der Informationsdimension 'Bedeutung&Verwendung' illustriert. Neben den bereits erwähnten kategorialen Einordnungsmöglichkeiten (z.B. 'Gib mir alle komparativen Phraseologismen'); lassen sich Wortverbindungen zunächst über unsere Lesartenetikettierungen (LesIDA) ordnen. Hierbei handelt es sich um eine Kurzformel zur Kennzeichnung von Lesarten der Lemmata, z.B.:

- I. 'Hund+Verb+begraben'
 <LesIDA> Kern eines Problems <LesIDA>
 II. 'Hund+Verb+begraben'
 <LesIDA> verlassener Ort <LesIDA>
- 'Nägel mit Köpfen'
 <LesIDA> erfolgreich handeln <LesIDA>

Man könnte dann alle Mehrworteinheiten herausfinden, die den Kern eines Problems, einen verlassenen Ort oder erfolgreiche bzw. ergebnisorientierte Handlungen bezeichnen.

Im Teilbereich der pragmatischen Angaben lassen sich neben diatopischen Zuordnungen ('peu à peu' = französisch, 'etw. aus Daffke tun' = berlinisch) u.a. folgende Felder bilden¹⁴:

- alle Mehrwortverbindungen, die eine textstrukturierende Funktion haben ('es war einmal' = Textanfang, 'zum Beispiel' = Textflussunterbrechung/Veranschaulichung);
- alle Mehrwortverbindungen, die an spezielle Textsorten gebunden sind bzw. selbst eine Textsorte konstituieren ('wie aus zuverlässiger Quelle verlautet' = Nachricht 'Wir sind das Volk' = Losung);
- alle Mehrwortverbindungen, die eine bestimmte Textfunktion indizieren ('Sehr geehrte Damen und Herren' = kontakktiv, 'an Eides statt' = deklarativ)
- alle Mehrwortverbindungen, die zu einer bestimmten Domäne gehören ('Wache schieben' = Militär: 'bei jmd. vorstellig werden' = Staat/Politik/Verwaltung/Recht)
- alle Mehrwortverbindungen, die einen fachsprachlichen Ausdruck darstellen ('freie Radikale' = Medizin)
- alle Mehrwortverbindungen, die eine spezifische Gruppenzugehörigkeit des Sprechers indizieren ('in medias res' = bildungsbürgerlich; 'voll/echt geil/cool' = jugendsprachlich)

¹³ Zu Thesaurus-Modellen und Taxonomien vgl. ausführlich Dobrovolskij (1995).

¹⁴ Viele der hier aufgeführten Beispiele sind in der Literatur vielfach besprochen worden (u.a. bei Palm 1995). Sie sind darüber hinaus – und das ist unser Kriterium – alle auch im Korpus belegt.

- alle Mehrwortverbindungen, die in spezifischen Kommunikationssituationen verwendet werden und ein spezielles sprachliches Register darstellen ('ein Bäuerchen machen' = Medialität mündlich, Register=vornehmlich in familiärer Atmosphäre)

5 Schlussbemerkung

Es sollte deutlich geworden sein, dass WiW nicht nur für den Nachschlagenden, sondern auch für den an wissenschaftlicher Auswertung interessierten Linguisten von Nutzen sein wird. Für die Idiomatikforschung kann dies perspektivisch nicht nur einen quantitativen Sprung bedeuten, indem ein bis dato in dieser Größe nicht existierender Fundus an aktuellen – also wirklich gebräuchlichen – Wortverbindungen des Deutschen bereitgestellt wird, sondern auch einen qualitativen Fortschritt. Idiomatisches Wissen kann empirisch verifiziert, expliziert und beschrieben werden. Es wird möglich, auch die ko- und kontextuellen Gebrauchsbedingungen für Mehrworteinheiten in einer ungleich systematischeren Weise einzubeziehen, als das bisher auch nur annähernd geleistet werden konnte.

Literatur

- Belica, C., 2002. *Die COSMAS-Kollokationsanalyse*. <http://www.ids-mannheim.de/kt/kollok.html>
- Boguraev, B. / J. Pustejovsky, 1996. *Corpus Processing for Lexical Acquisition* [= Language, Speech, and Communication]. Cambridge, MA.
- Burger, H. (1998), *Phraseologie. Eine Einführung am Beispiel des Deutschen* [= Grundlagen der Germanistik 36]. Berlin.
- Dobrovol'skij, D., 1993. "Datenbank deutscher Idiome. Aufbauprinzipien und Einsatzmöglichkeiten". In: Földes, C. (Hrsg.). *Germanistik und Deutschlehrerausbildung*. Szeged / Wien, 51-67.
- Dobrovol'skij, D., 1995. *Kognitive Aspekte der Idiom-Semantik. Studien zum Thesaurus deutscher Idiome* [= Eurogermanistik 8]. Tübingen.
- Feilke, H., 1996. *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt a. M.
- Feilke, H., 2000. "Textroutine, Textsemantik und sprachliches Wissen". Bielefeld [Vortragsmanuskript].
- Fleischer, W., 1997. *Phraseologie der deutschen Gegenwartssprache*. 2., durchgesehene u. ergänzte Aufl. Tübingen.
- Fraas, C. / U. Haß-Zumkehr, 1998. "Vom Wörterbuch zum lexikalischen Informationssystem LEXXIS – ein neues Projekt des Instituts für deutsche Sprache". In: *Deutsche Sprache* 26, 289-303.
- Gülich E. / U. Krafft, 1998. "Zur Rolle des Vorgeformten in Textproduktionsprozessen". In: Wirrer, J. (Hrsg.). *Phraseologismen in Text und Kontext* [= Phrasemata I]. Bielefeld, 11-38.

- Hausmann, F. J., 1984. "Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen". In: *Praxis des neusprachlichen Unterrichts* 31, 395-406.
- Haß-Zumkehr, U., 2001. "Zur Mikrostruktur im Hypertext-Wörterbuch". In: Lemberg, I. / B. Schröder / A. Storrer (Hrsg.). *Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*. Tübingen [erscheint].
- Haß-Zumkehr, U., 2002. "Das Wort in der Korpuslinguistik. Chancen und Probleme empirischer Lexikologie". In: Agel, V. / A. Gardt / U. Haß-Zumkehr / T. Roelcke (Hrsg.). *Das Wort ... Festschrift für Oskar Reichmann zum 65. Geburtstag*. Tübingen [in Vorber.].
- Heringer, H. J., 1999. *Das höchste der Gefühle. Empirische Studien zur distributiven Semantik* [= Stauffenburg-Linguistik]. Tübingen.
- Lemnitzer, L., 1997. *Akquisition komplexer Lexeme aus Textkorpora* (= Germanistische Linguistik 180). Tübingen.
- Palm, C., 1995. *Phraseologie. Eine Einführung*, Tübingen.
- von Polenz, P., 1985. *Deutsche Satzsemantik. Grundbegriffe des Zwischen-den-Zeilen-Lesens* [= Sammlung Götschen 2226]. Berlin / New York.
- Steyer, K., 1997. *Reformulierungen. Sprachliche Relationen zwischen Äußerungen und Texten im öffentlichen Diskurs* [=Studien zur deutschen Sprache 7]. Tübingen.
- Steyer, K., 1998. "Kollokationen als zentrales Übersetzungsproblem – Vorschläge für eine Kollokationsdatenbank Deutsch-Französisch / Französisch-Deutsch auf der Basis paralleler und vergleichbarer Korpora.". In: Bresson, D. (Hrsg.). *Lexikologie und Lexikographie Deutsch-Französisch* [= Cahiers d'Études Germaniques 35]. Aix-en-Provence, 95-113.
- Steyer, K., 2000. "Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikografische Möglichkeiten". In: *Deutsche Sprache* 2/00, 101-125.
- Steyer, K., 2003. "Korpus, Statistik, Kookkurenz. Lässt sich Idiomatisches 'berechnen'?" In: H. Burger / A. Häcki Buhofer / G. Gréciano (Hrsg.). *Flut von Texten – Vielfalt der Kulturen. Ascona 2001 zu neuen Fragen und Methoden der Phraseologie*. Baltmannsweiler [im Druck].
- Stubbs, M., 1997. "'Eine Sprache idiomatisch sprechen': Computer, Korpora, Kommunikative Kompetenz und Kultur". In: Mattheier, K. J. (Hrsg.). *Norm und Variation* [= Forum Angewandte Linguistik 32]. Frankfurt a.M. / Berlin / Bern / New York / Paris / Wien, 151-167.

Wörterbücher

- DUW = DUDEN Deutsches Universalwörterbuch, 1996. 3., neu bearbeitete und erweiterte Aufl. Auf der Grundlage der neuen amtlichen Rechtschreibregeln. Bearb. von Günther Drosdowski und der Dudenredaktion, Mannheim / Leipzig / Wien / Zürich.
- DUDEN 11 = *DUDEN Redewendungen und sprichwörtliche Redensarten*, 1998. *Wörterbuch der deutschen Idiomatik*. Bearb. von Günther Drosdowski u. Werner Scholze-Stubenrecht. Nach den Regeln der neuen deutschen Rechtschreibung, überarb. Nachdruck der 1. Aufl. Mannheim / Leipzig / Wien / Berlin.
- Schemann, H., 1993. *PONS Deutsche Idiomatik. Die deutschen Redewendungen im Kontext*. Stuttgart / Dresden.